Technology at the meeting point of hardware, software and mindware

Bernard Bel

Technology at the Meeting Point of Hardware, Software and 'Mindware'

Bernard Bel

Laboratoire Parole et Langage (LPL)
UMR 6057, CNRS / Aix-Marseille Univ
5, avenue Pasteur
13100 Aix-en-Provence (France)
bernard.bel@lpl-aix.fr

Abstract

The aim of this paper is to reflect upon the impact of technology on data collection, analysis and interpretation in a work environment for the sharing of research material, tools and methods by experimental and field linguists. An analogy is drawn between expectations of this approach and successful contributions of experimental/computational techniques to ethnomusicology in the past decades.
The author is in charge of the development and maintenance of the *Speech & Language Data Repository* (SLDR), a submission site on the OAIS model for the long-term preservation of linguistic resources. Items in the repository are generic, which makes it possible to extend the notion of corpus to all kinds of primary data collected during experimental or field sessions. This flexibility has an incidence on the life cycle of research projects, as exemplified by an 'event-driven' project on the documentation and revitalisation of rare local languages conducted in narrow collaboration with the development of the archival system.

Background

A research engineer in a multidisciplinary speech research laboratory (*Laboratoire Parole et Langage*),[1] the author is in charge of the development of the *Speech & Language Data Repository* (SLDR),[2] a free-of-charge service for sharing oral/linguistic data and archiving it with the help of procedures compliant with the OAIS model for long-term preservation.[3] This service has been implemented in the context of a pilot project coordinated by TGE-Adonis in 2008-2010.[4]

An innovative aspect of SLDR is its ability to cope with the life cycle of research projects from the initial description of resources in the making to their long-term preservation in an institutional archive.[5] This implies a trade-off between the versatility of data/metadata — upgrading versions and setting up their access rights — and legal constraints associated with the use of a public archive. Access rights must stay compliant with the French heritage code (*Code du patrimoine*) which has provision for retention periods with respect to personal data (Bel, 2011).

In order to assess the relevance of implemented features, priority was given to checking in great detail the OAIS model on material produced by a small number of projects in very different domains. In addition, new projects have been undertaken along parallel lines with the development of SLDR, thereby raising issues on methodology in terms of technical performance as well as interactions between archive curators, research scholars and project participants/informants. In this context, a project dealing with the documentation and revitalisation of a 'forgotten' local language in the North Occitan area (*patois de Valjouffrey*)[6] was launched in 2009 under the direction of sociolinguist Médéric Gasquet-Cyrus. The author is involved as a collaborator and sound engineer. The project team is now in charge of a cross-disciplinary work group dealing with cross-fertilisation (and contradictions) between experimental and field approaches in linguistics.[7]

Early investigations in ethnomusicology

To some extent, the Valjouffrey project and, more generally, linguistic resource collection and analysis projects taking advantage of new technology, are reminiscent of issues raised by ethnomusicology in the early 1980s. In those days, new tools and methods were being developed for the study of productions and interactions in musical performance and training. The main targets of this investigation were the tonal and rhythmic systems of North Indian classical music, the former of which had been a subject of speculative debate for more than two centuries (Bel, 1988; 1992). The challenge of a 'hard-science' study was to reconcile the approach of systematic musicology — looking at music as *material* lending itself to structural analysis — with that of musicology centred on *musicians* and *processes* of creation/reception/institutionalisation (Bel & Bor, 1984; Bel & Vecchione, 1992).

Significant results had been achieved owing to the combination of two methodologies: (a) collecting the largest possible amount of data from musical performances in a format suitable for statistical analysis, and (b) constructing a computational model of music production/perception reflecting human processes and evaluated by its own informants, i.e. inductive learning

[1] http://lpl-aix.fr

[2] http://sldr.org

[3] Open Archival Information System, http://public.ccsds.org/publications/archive/650x0b1.pdf

[4] TGE-Adonis, http://tge-adonis.fr. The set of working papers and reports produced during the pilot project is available from a public archive: http://sldr.org/ark:/87895/1.4-187408

[5] *Centre Informatique National de l'Enseignement Supérieur* (CINES), http://cines.fr

[6] http://sldr.org/wiki/Valjouffrey

[7] *Groupe de recherche transversal « Méthodologies linguistiques »* in Laboratoire Parole et Langage http://ling-metho.hypotheses.org

techniques associated with an expert system[8] (Bel & Kippen, 1989).

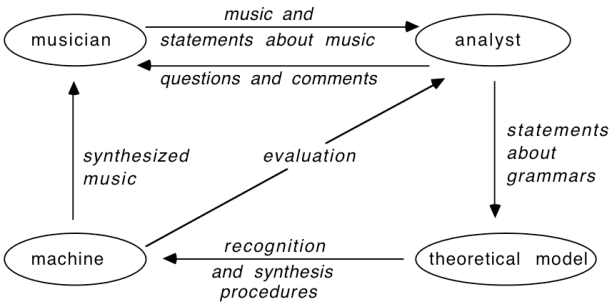

Figure 1: A flowchart of interactions in an expert-system setup for the identification of drum improvisation 'languages' in North India (Kippen & Bel, 1989).

Analysis of oral poetry

In 1995 the author became aware of a monumental documentation work conducted for several decades in rural Maharashtra by a group of social activists and scholars under the banner of *Centre for Cooperative Research in Social Science* (CCRSS)[9] and *Village Community Development Association* (VCDA).[10] In the context of action-research projects aiming at the emancipation of down-trodden communities, the team was addressing forms of expression and communication displaying a characteristic continuity between 'traditional' and 'modern' thought processes. Among these, social animators had been involved in the collection of songs composed and performed by peasant women at dawn during their work at the grindmill.

At this stage of the project, more than 100,000 song texts have been transcribed (in Marathi *devanāgarī* script) and their contents analyzed in the framework of a detailed semantic classification scheme. This corpus has been stored in a relational database (see figure 2) linking texts with information on performers, villages, pictures and an index of more than 100 hours of sound recordings highlighting their specific modes of performance (Bel, Caelen-Haumont & Rairkar, 2000). The entire corpus is in the process of being shared in open access along with English translations of texts and all sound fragments listed in the index.[11]

This knowledge base became fully operational after the implementation of a LEXICON table containing lexemes and words from which a script can be activated to compile occurrences of their written variants in the SONGS database and a profile of these occurrences with respect to spatial location (districts and *talukas*), periods of time and the social status of performers. This query tool was used extensively by Guy Poitevin and Hema Rairkar to capture semantic and contextual information about concepts specific to the oral poetic language shared by peasant women of all classes. This process facilitated the work of extracting thematic corpora[12] and using them for publications (Poitevin & Rairkar, 2008; Bel, ed., 2009).

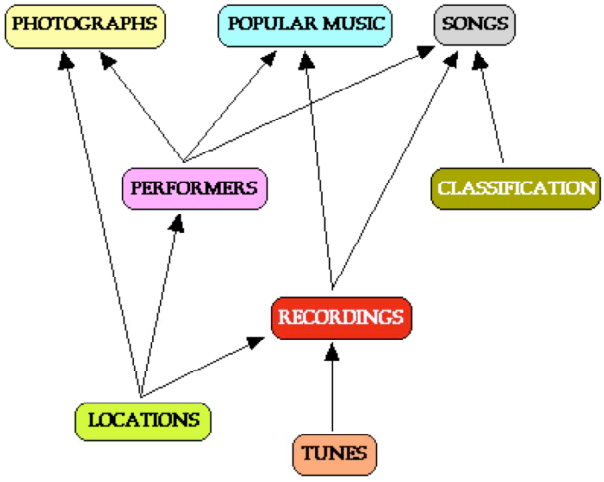

Figure 2: The relational database of grindmill songs at CCRSS in 1996.

New trends

The notion of 'corpus'

Designing SLDR as a *generic* submission site for an OAIS archive led us to redefine a 'corpus' as a non-restrictive set of *primary data* collected during experimental or field sessions. In experimental linguistics, corpora may include audio/video recordings as well as physiological activity measurements: endoscopy, evoked potentials, palatography, etc. In field sessions video recording is becoming systematic for a better tracking of interactions and growing interest in the analysis of gestures and facial expressions. Sound recordings are of higher quality thanks to head-worn microphones recorded on multiple separate tracks.

In order to preserve and share 'historical' corpora (such as EUROM[13]) it was necessary to accept that the information package of an information package in category 'primary data' may include secondary data (annotations etc.).

In our field research, all participants have been incited to hand over informative documents from various sources which in the long term might become indispensable for a proper recollection of contextual information associated with interactions and interviews: photographs, drawings, written notes, maps, computer documents etc.

Our broad interest in primary data is consistent with the idea of collecting "the largest possible amount of data" which had been successful in ethnomusicology. Indeed, it may look extravagant to store more than 150 Gbytes of sound/video recordings, photographs and written documents in a single *Valjouffrey 2010-2011 corpus*.[14] Critical minds are asking what *we* plan to do with *all* this data… There are at least two answers:

---

[8] *Bol Processor*. oai:sldr.org:sldr000753

[9] http://ccrss.ws

[10] http://vcda.ws

[11] oai:sldr.org:sldr000717, oai:sldr.org:sldr000735. Recordings are also available with the *Archive and Research Center for Ethnomusicology* (ARCE), New Delhi.

[12] *Gangubai*, oai:sldr.org:sldr000759. *Ambedkar*, oai:sldr.org:sldr000011

[13] oai:sldr.org:sldr000741

[14] oai:sldr.org:sldr000764

1) The data we collect is meant to be shared with the research community at large. Different teams from a wide range of disciplines might get involved: descriptive linguistics, sociolinguistics, linguistic anthropology, social psychology and history, to name a few.
2) The agenda of future research will focus on pieces of information which the original team had judged merely contextual. For instance, the scans of hand-written municipal archives may be appended to the corpus as they are required for drawing genealogic trees. This suggestion is supported by the fact that factual information on lineage is a sensitive topic of discussion in the community under study.
Our claim to collect, preserve and share exhaustive research material beyond the range of linguistics deserves criticism for its lack of methodological insights. An anthropologist or an historian may feel worried that non-directive interactions eluded 'the right questions'; an experimental linguist may claim that a careful planning of lab experiments would highlight language features which an uncontrolled approach is likely to miss (Xu, 2010). All this criticism is relevant though it should be balanced against the patrimonial value of a corpus dealing with rare or endangered languages (Grenoble & Whaley, 2006; Austin & Sallabank, 2011).

Versatile knowledge

We had a debate in the TGE-Adonis pilot project as to whether items in a repository should only contain atomic entities in a fixed format (e.g. a single sound file) or an unrestricted collection of digital documents in multiple formats. Following the extended notion of 'corpus' we (at SLDR) decided to implement a model handling generic information packages. These are automatically segmented (in a manner invisible to users) when the limits of 30,000 files and/or 40 Gbytes imposed by the archive and dissemination sites is reached. File names and hierarchical tree structures are re-encoded in such a way that any complex structure with 'exotic' names (in Unicode format) is acceptable, and downloaded items are reconstructed identical to the source.[15]
Maintaining 'heavy' packages in a long-term preservation scheme is problematic since all versions must be preserved by the archive. This led us to divide each item into two forks: a DATA fork containing stable documents (e.g. source recordings and photographs) and a DESC fork containing a set of descriptive documents that may need to be upgraded from time to time. In practice, the DESC fork does not contain primary data, which amounts to calling it 'metadata' in a broad sense.

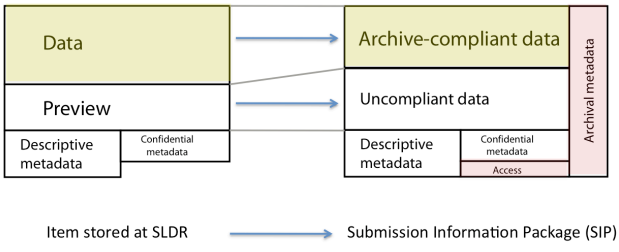

Figure 3: The making of a submission information package on SLDR

Our implementation of the OAIS model has made it possible to submit new versions of descriptive files (a *metadata update*) in relation with the same version of data files. This makes it unnecessary to upload the entire corpus each time metadata or descriptive files have been modified.
Further, we implemented a management of access rights that makes it possible to modify access attributes via a simple update of descriptive files.[16]
This flexible versioning technique has tremendous implications on the phasing of projects associated with the production of linguistic resources:
1) Scholars create records describing the data and resources they plan to elaborate during the project. These records may remain invisible if they wish so. They provide persistent URIs that may be cited in project proposals.
2) Data collected during field or experimental sessions is uploaded to the corpus and stored in medium-term preservation. This type of storage produces versions that may be deleted once the item has been submitted for long-term preservation.
3) Storage in medium-term preservation is identical to storage in long-term preservation with respect to the dissemination of data and metadata (the OAI-PMH repository). Thus it is possible for project teams to immediately start sharing selected parts of their corpus and its annotations. Sharing may be open-access or restricted based on user/institution groups or more restrictive targets authenticated by identifiers and passwords. Thus, medium-term preservation is a period of intensive activity during which data producers may take advantage of sharing partial results with fellow members of the work group as well as other teams and individuals.
4) Once the data contained in a package has become stable, it may be submitted for long-term preservation. Former versions are deleted and a new thread of versioning is started. Since all versions must there from be preserved, submitting a new version should be motivated. Nonetheless, submitting new versions of descriptive files is a routine procedure. This makes it possible to modify access rights once intellectual property issues have been sorted out and informed consent documents signed by participants (with their scans uploaded to the archive as 'descriptive files'). Further, restricting access (to the current and previous versions) of a document is technically easy in case participants withdraw their consent, which must be possible at any time according to the Law.
In short, SLDR management makes it possible to modify access rights to documents despite their immutable content in the archive. This flexibility is a significant advance given that failure to collect the full set of informed consent documents before the completion of a project is often dissuasive of handing over its corpus to an institutional archive.

Event-driven field research

Our fieldwork on Valjouffrey and neighbouring dialects may be called 'event-driven' rather than 'protocol-driven'. Language description was not the focus of our work because a priority had been given to creating conditions in

[15] See SLDR *Packaging items*, http://sldr.org/wiki/Packaging-en

[16] http://sldr.org/wiki/accessRightsSettings_en

which the very small number of speakers (aged from 65 to 82) would feel at ease to communicate and revive a language they had not practiced for two decades. The initial impulse was therefore antagonist with linguistic purism, codification and the dream of reviving 'authentic' or 'traditional' language usage.

Despite this, as we shared documents on SLDR wiki pages[17] cited by a local non-profit society interested in the valley's cultural heritage, we received calls and mails from two retired experts, Clément Girard and Marcelle Péry, who had produced unpublished dissertations on the Valjouffrey[18] and Valbonnais[19] *patois* respectively. They quickly became members of our team.

In 2010-2011, Marcelle Péry organised regional encounters on patois at her home in Valbonnais. These events have been entirely covered in video for public access.[20]

In Summer 2010, Julien Gaillard, our senior informant in Valjouffrey, decided on his own to document the names of places and details relevant to mountain climbing in the valley. To this effect he sought assistance from Robert Jamos, an artist and mountain-climber, for the drawing of accurate views of the documented sites. This event oriented our research to the documentation of verbal interactions associated with the elaboration of this toponymy.[21]

Event-driven research produces information packages of very diverse contents. Hand-written dissertations have been scanned, archived and disseminated in open access. More than ten hours of old recordings supplied by Clément Girard have been digitized and are available in open access as a complement of his scanned dissertation. Drawings and maps, lexica will be preserved as 'linguistic resources' attached to the corpus, etc. This is further evidence that an archive submission site should be able to manage generic items.

Figure 4: A recording session in Valjouffrey, using head-worn microphones and a multitrack recorder.

'Empowering' linguistic research

In the research community at large, i.e. beyond institutional and geographical borders, there is a growing urgency to build distributed research infrastructures of integrated and interoperable language resources and tools that serve researchers and the students in the humanities and social sciences (Wittenburg, 2010). In Summer 2011 partners of the ORTOLANG project[22] decided to draw from the experience of SLDR and CNRTL (*Centre National de Ressources Textuelles et Lexicales*)[23] to build a network infrastructure offering a repository of language data (corpora, lexicons, dictionaries, etc) and tools and their treatment, set up according to the guidelines of CLARIN centres.[24] Among arguments of this proposal is the contention that *"[…] in our society of information, only the strongly equipped and modelled languages that can be used for automatic language processing, are likely to remain as common languages for work and exchange in scientific, economic, industrial and cultural disciplines. The remaining ones are likely to be reduced to their simple vernacular dimension."*

To meet this challenge, simple descriptive linguistics is being superseded by formal linguistics covering lexical, syntactical or semantic aspects, working with models based on a double validation, *explanatory* from a linguistics point of view and *operational* from a data-processing point of view. This has given way to the emergence of *corpus linguistics* (Habert, Nazarenko & Salem, 1997) making it possible for linguists to go beyond the accumulation of language facts and to confront their theories with the effective use of language.

This change of paradigm has strong implications both on the production of vast linguistic resources of high quality (corpora, dictionaries and lexicons) and the structuring and standardisation of linguistic knowledge — phonetics, morphology, lexical, syntactics, semantics, etc. — imbedded in annotations and metadata.

Annotation

The issue of annotation is an important focus of debate and experimentation in large-scale projects (such as OTIM[25]) notably because of a growing interest in interaction processes, the formal study of which requires taking into account the whole set of modalities and the way they interact. This will render experimental linguistics increasingly attractive (and challengeable) to field linguists who have long been associated with interactional analysis (Gumperz, 1982; Mondada & Markaki eds., 2006).

When dealing with small groups of speakers, work is made easier owing to multitrack recordings of head-worn microphones. We are now working on the automatic pre-processing of separate tracks that will identify the borders of phrase units, perform basic prosodic analysis (MOMEL/INTSINT)[26] and create a stereo mix using panoramic/volume setups completed with a PRAAT TextGrid combining annotations tiers for all speakers.

Multimodal information may be captured with the aid of standard video recording, a technique that is increasingly

[17] http://sldr.org/wiki/Valjouffrey

[18] oai:sldr.org:sldr000006

[19] oai:sldr.org:sldr000005

[20] oai:sldr.org:sldr000736

[21] Video work session (13 February 2011): P64m1, P64m2, in oai:sldr.org:sldr000764

[22] http://sldr.org/wiki/ORTOLANG

[23] http://www.cnrtl.fr

[24] http://www.clarin.eu

[25] OTIM - *Tools for Multimodal Information Processing*, http://lpl-aix.fr/~otim

[26] MOMEL, oai:sldr.org:sldr000031. INTSINT, oai:sldr.org:sldr000032.

affordable, including 3D which produces material for future automatic gesture/expression analysis. In addition, the technologies of articulatory phonology are bound to become accessible for field research, e.g. the evaluation of speech production (including airflow and pressure),[27] ultrasound tongue imaging[28] and (to some extent) electromagnetic articulography (Kaburagi & Honda, 2002).

Figure 5: Electromagnetic articulography in a laboratory setup (*Centre d'Expérimentation sur la Parole* at LPL).

Standardisation

Standardisation is a technical obligation for data repositories playing the game of interoperability. This should not be achieved at the cost of becoming unable to cope with linguistic diversity. As shown by variationist sociolinguists the complexity of real-life communication processes may in term become a subject of systematic research in which regularities are traced at a different level and scale (Tagliamonte, 2011; Labov, 1994, 2001, 2010).

We must also keep in mind that standardizing the symbols that make sense to speakers/writers of a local language may be problematic because of the associated loss of cultural identity at the benefit of scholars perceived as 'outsiders'. *"Distinct scripts support notions of independence"* (Spolsky 2004: 29). Nonetheless, the same local actors feel the need to resort to normalisation as it reinforces their roles as experts in the context of language revitalisation (Thomas-Aguillon, Gasquet-Cyrus & Bel, 2011).

The interest of a formal/computational approach is that background knowledge is made explicit by the training of inductive-inference devices (Bel & Kippen, 1989). This makes it possible to displace the focus of research from the standard description of a linguistic/musical process to questioning the validity of 'background' information mobilised for this description. Individual and group variations are taken into consideration after having been neglected in the process of standardisation.

Keeping in mind this care for variation and diversity, research scholars need to elaborate data categories and associated value domains to be shared by humans and machines, for which a sophisticated representation framework is required. Such is the challenge of ISOcat.[29]

Informants turn researchers

When trying to identify drum improvisation 'languages' in an expert-system setup in the early 1980s, the author was inspired by a methodology promoted by anthropologist/musicologist John Blacking at the Queen's University of Belfast: let informants participate in both the *elaboration* and the *validation* of models describing their expertise.[30] Applied to the grammars of finite languages, this approach lent itself to computational interactions that rendered language identification feasible in an acceptable number of steps (Kippen & Bel, 1992).

Dealing with infinite (*natural*) languages is not so straightforward. It remains that if a certain amount of expertise can be embedded in a computational model, expert speakers/writers will take interest and may eventually feel proud of its consistency. This is one further step in an empowerment starting with assertions of a normative symbolic system (script and spelling).

Figure 6: Audrey Thomas-Aguillon, Robert Bois and Hubert Balmet during a work session for the elaboration of a script of *patois de Valjouffrey*

The quest for consistency is an incentive for experts to participate in experiments based on sophisticated protocols, for instance the AMPER study of intonation patterns of Roman languages,[31] designing a script for their

[27] EVA2, http://www.sqlab.fr/evaRootUK.htm

[28] SSRC Ultrasound Tongue Imaging, http://www.qmu.ac.uk/ssrc/ultra/

[29] ISOcat Data Category Registry, ISO 12620:2009, http://www.isocat.org/files/12620.html

[30] Though Blacking suggested the term 'dialectical anthropology' to designate this approach, this term is now used with a different denotation, see http://en.wikipedia.org/wiki/Dialectical_Anthropology

[31] AMPER : Atlas Multimédia Prosodique de l'Espace Roman, http://w3.u-grenoble3.fr/dialecto/AMPER/amper.htm

revitalized language,[32] or undertaking a detailed inventory of place names that delineate their living space.

Conclusion

Our interdisciplinary approach aims at reconciling (a) the quality requirements of experimental linguistics, (b) needs for the preservation and pooling of resources and (c) ethics and quality requirements of field linguistics.

The material versus human and data versus process controversies are always reactivated when technology makes it possible to deal with accurate statistical measurements on a population of presumably similar human subjects (as it is the case in neurosciences) while field research remains cautious of ethnocentrism and individual variation. Thus, musicologists using computational methods are still on a knife edge, and unsurprisingly similar questions are raised by linguists addressing the study of speech and communication processes at large.

The ambitious agenda of event-driven field research is to capture and theorize *"the motivations and perceptions of speech communities"* beyond the initial objective of describing their languages *"address[ing] a narrow audience of 'external' linguists with similar concerns…"*[33]

To achieve this, equipment (*hardware*) and methods (*software*) confined in laboratories need to be adapted to their usage in field sessions trying to capture parameters which so far had been studied qualitatively or ignored by linguists.

Success in this direction will depend on the capacity of research scholars to share data, acquisition and analysis techniques, combinations of qualitative and quantitative approaches, interpretation of results, and insights into innovative methodologies — all of which may be termed as *'mindware'*. This porosity of research domains is among the major challenges of our *Laboratoire Parole et Langage* and its unique CEP experimental platform.[34]

Acknowledgements

This paper is the outcome of passionate discussions with colleagues at LPL and more particularly during our fieldwork on the Valjouffrey project. Médéric Gasquet-Cyrus suggested most recent bibliographical sources.

Bibliographical References

Austin, P.K.; Sallabank, J. (2011). The Cambridge Handbook of Endangered Languages. Cambridge (UK): Cambridge University Press.

Bel, B. (1988). *Raga: approches conceptuelles et expérimentales.* In Proceedings of *Colloque Structures Musicales et Assistance Informatique* (pp. 87--108). Marseille, France. http://hal.archives-ouvertes.fr/hal-00008280

Bel, B. (1992). Modelling improvisatory and compositional processes. Languages of Design, Formalisms for Word, Image and Sound, 1(1), 11--26. http://hal.archives-ouvertes.fr/hal-00256385

Bel, B. (Ed.) (2009). *Ambedkar ! Des Intouchables chantent leur libérateur. Texte de Guy Poitevin édité à titre posthume.* ISBN 978-2-8111-0216-6. Paris: Karthala http://hal.archives-ouvertes.fr/hal-00005792

Bel, B. (2011). Access rights management in compliance with the French *Code du patrimoine*: a generic approach for the OAIS model run by SLDR. Online document. http://sldr.org/wiki/accessRightsManagement_en

Bel, B.; Bor, J. (1984). Intonation of North Indian Classical Music: working with the MMA. Video documentary. Mumbai (India): National Centre for the Performing Arts. http://lpl-aix.fr/article/4820

Bel, B.; Kippen, J. (1989). The identification and modelling of a "percussion" language, and the emergence of musical concepts in a machine-learning experimental set-up. Computers and the Humanities, 23(3), 199--214.

Bel, B. ; Vecchione, B. (eds.) (1992). Music Sciences and Technologies: New Inquiries, New Connections. Interface, Journal of New Music Research, 21, 3--4. http://halshs.archives-ouvertes.fr/halshs-00004505

Bel, B.; Caelen-Haumont, G.; Rairkar, H. (2000). Prosodic Patterns in the Performance of Grindmill Songs. International Seminar on Linguistic and Inter-Disciplinary Approaches as Critical Resources to Development. Mysore, India, 12-14 July. http://hal.archives-ouvertes.fr/hal-00256387

Grenoble, L.A. & Whaley, L.J. (2006). Saving languages: an introduction to language revitalization. Cambridge (UK): Cambridge University Press.

Gumperz, J. (1982). Discourse Strategies (Studies in Interactional Sociolinguistics). Cambridge (UK): Cambridge University Press.

Habert, B., Nazarenko, A. et Salem, A. (1997). *Les linguistiques de corpus.* Paris: Armand Colin.

Kaburagi, T.; Honda, M. (2002). Electromagnetic articulograph based on a nonparametric representation of the magnetic field. J. Acoust. Soc. Am., 111(3), 1414--1421.

Kippen, J.; Bel, B. (1989). Can a computer help resolve the problem of ethnographic description? Anthropological Quarterly, 62(3), 131--144. http://hal.archives-ouvertes.fr/hal-00275429

Kippen, J.; Bel, B. (1992). Modelling music with grammars: formal language representation in the Bol Processor. In A. Marsden & A. Pople (Eds.) Computer Representations and Models in Music (pp. 207--238). London: Academic Press. http://halshs.archives-ouvertes.fr/halshs-00004506

Labov, W. (1994, 2001, 2010). Principles of Linguistic Change (3 volumes). Hoboken, NJ: Wiley-Blackwell.

Mondada, L.; Markaki, V. (Eds.) (2006). Interacting Bodies. Proceedings of the 2d ISGS Conference. http://gesture-lyon2005.ens-lsh.fr/article.php3?id_article=259

Poitevin, G.; Rairkar, H. (2008). Bhakti, a Faith for Rehabilitation. In B. Bel, J. Brouwer, B. Das, V. Parthasarathi, G. Poitevin (Eds.) Communication

[32] Audio/video work sessions: P38-39-40-41, 29 June 2010, and P59-P60, 12 February 2011, oai:sldr.org:sldr000764

[33] Cited from project *Endangered language revitalisation: theory, background and implementation* submitted to the Open Research Area in Europe for the Social Sciences in October 2011

[34] *Centre d'Expérimentation sur la Parole*, http://lpl-aix.fr/~cep/indexa.html

Processes, 2: The Social and the Symbolic (pp. 170--215). New Delhi: Sage. http://sldr.org/doc/000759/BhaktiAFaithForRehabilitation.pdf

Spolsky, B. (2004). Language policy. Cambridge (UK): Cambridge University Press.

Tagliamonte, S.A. (2011). Variationist Sociolinguistics: Change, Observation, Interpretation. Hoboken, NJ: Wiley-Blackwell.

Thomas-Aguillon, A.; Gasquet-Cyrus, M.; Bel, B. (2011). *Revitalisation ou reniement de la langue locale ? L'élaboration problématique d'une graphie pour le « patois » de Valjouffrey*. Colloque international AULF et LESCLaP-CEP: *Standardisation et vitalité des langues de France*. October 13-14, Amiens, France. http://lpl-aix.fr/article/4712 (in press)

Wittenburg, P. (2010). A presentation of CLARIN. Online document. http://www.clarin.gr/files/wittenburg%20presentation-athens.pdf

Xu, Yi (2010). In defense of lab speech. Letter to the Editor. Journal of Phonetics, 38, 329-–336. http://www.phon.ucl.ac.uk/home/yi/yispapers/Xu_JPhon2010_author_version.pdf